AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Human Perception and Performance

ISSN: 0096-1523

2023, Vol. 49, No. 6, 774–785
https://doi.org/10.1037/xhp0001118

RESEARCH REPORT

Your Ears Don't Change What Your Eyes Like: People Can Independently Report the Pleasure of Music and Images

Jessica Frame[1], Maria Gugliano[2], Elena Bai[1], Aenne Brielmann[3], and Amy M. Belfi[1]
[1] Department of Psychological Science, Missouri University of Science and Technology
[2] College of Pharmacy, University of Texas at Austin
[3] Department of Computational Neuroscience, Max-Planck Institute for Biological Cybernetics

Observers can make independent aesthetic judgments of at least two images presented briefly and simultaneously. However, it is unknown whether this is the case for two stimuli of different sensory modalities. Here, we investigated whether individuals can judge auditory and visual stimuli independently, and whether stimulus duration influences such judgments. Participants ($N = 120$, across two experiments and a replication) saw images of paintings and heard excerpts of music, presented simultaneously for 2 s (Experiment 1) or 5 s (Experiment 2). After the stimuli were presented, participants rated how much pleasure they felt from the stimulus (music, image, or combined pleasure of both, depending on which was cued) on a 9-point scale. Finally, participants completed a baseline rating block where they rated each stimulus in isolation. We used the baseline ratings to predict ratings of audiovisual presentations. Across both experiments, the root mean square errors (RMSEs) obtained from leave-one-out-cross-validation analyses showed that people's ratings of music and images were unbiased by the simultaneously presented other stimulus, and ratings of both were best described as the arithmetic mean of the ratings from the individual presentations at the end of the experiment. This pattern of results replicates previous findings on simultaneously presented images, indicating that participants can ignore the pleasure of an irrelevant stimulus regardless of the sensory modality and duration of stimulus presentation.

Public Significance Statement

This study suggests that people are able to make independent judgments of stimuli in their environment, for example, that a person can accurately judge how much they like a painting even if listening to a song they dislike.

Keywords: music, aesthetics, pleasure, reward

Supplemental materials: https://doi.org/10.1037/xhp0001118.supp

During everyday life, individuals regularly encounter multiple, simultaneous sources of sensory information. As an example, consider visiting a restaurant. While eating their meal, a diner is also exposed to sights (e.g., artwork, decor), sounds (e.g., conversations, music), and smells (e.g., other foods). Someone might walk away with a negative impression of a particular dish, but rather than being a "pure" judgment of the food itself, they may have been influenced by these other stimuli. In contrast to this naturalistic context in which people are exposed to multiple stimuli, most prior work has taken a reductionist approach to investigating aesthetic judgments. That is, participants are typically asked to make judgments about items in a single sensory modality while eliminating all other sources of sensory input, such as judging an image of a single painting in isolation. Here, we sought to answer the question of whether individuals can make an independent aesthetic judgment of an item while simultaneously exposed to a stimulus in a different modality. To give another real-world example: If a person views their favorite painting while listening to a song they detest, do their negative feelings about the music influence their judgment of the painting?

This article was published Online First May 4, 2023.

Amy M. Belfi https://orcid.org/0000-0002-1469-4117

We would like to thank Amanda Lim and Mikayla Shy for their assistance with conducting literature reviews and setting up the experiment. Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Number R15AG075609. Aenne Brielmann was supported by the Max Planck Society, the Alexander von Humboldt Foundation, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–461354985.

Raw data, analysis files, and details of all packages used are available at https://github.com/aenneb/image_and_music_pleasure. This study was not preregistered.

Correspondence concerning this article should be addressed to Amy M. Belfi, Department of Psychological Science, Missouri University of Science and Technology, 135 H-SS, 500 W. 14th Street, Rolla, MO 65409, United States. Email: amybelfi@mst.edu

Judgments of Simultaneously Presented Images

Recent work sought to first address the question of whether observers can independently judge the pleasure of two stimuli presented in the *same* sensory modality. Researchers briefly presented two images side-by-side and observers were asked to rate the pleasure of one of the images (Brielmann & Pelli, 2020). Based on predictions from prior literature, the authors tested multiple models for how observers might rate the pleasure of a single image in the presence of a second, distractor image. Additionally, the authors sought to investigate what happens when observers are asked to rate the combined pleasure of both images. They found that in both cases, a "faithful" model best accounted for the data. That is, observers could both faithfully report the pleasure of a single image when accompanied by a second, irrelevant image, and also faithfully report the average of both images when asked to do so.

These findings (Brielmann & Pelli, 2020) stand in contrast to a line of work which demonstrates that people's ratings of the perceptual properties of a single item among several are biased toward the average of the set (Brady & Alvarez, 2011; Haberman & Whitney, 2009; Maule et al., 2014). They thus suggest that a *subjective evaluation* of an object might be different from reporting an object's perceptual properties. That is, subjective ratings (i.e., ratings of pleasure in this case) of a single item seem less likely to be biased by other concurrent items in an array. However, it is not entirely clear whether pleasure judgments of multiple objects are unbiased under all circumstances. In a follow-up study designed to test the limits in terms of the number of concurrent stimuli, Brielmann and Pelli (2021) repeated the same task but with four simultaneously presented images. The authors found that participants could *not* faithfully report the pleasure of a single image unless cued to which image they would be rating ahead of time, and that observers also did not faithfully report the average pleasure of four images (Brielmann & Pelli, 2021). With longer presentation durations of 1.5 s (vs. 200 ms in Brielmann & Pelli, 2020) and a homogeneous image set, pleasure ratings of one scene among three others were found to be biased toward the mean pleasure of the presented scenes (Alwis & Haberman, 2020).

Thus, there seem to be clear limits to people's ability to independently report the pleasure of multiple objects that they encounter simultaneously—when presented with two images, observers faithfully report the pleasure of a single image, but not when presented with four images, or when images are presented for a longer duration. So far, these limits have only been tested within the visual modality. However, another important question is whether people can encode and report the pleasure of two objects of different modalities that are presented at the same time. In the current study, we sought to investigate whether pleasure judgments of simultaneously presented music and images are combined when asked to rate the image, the music, and the combination of both.

Judgments of Simultaneously Presented Images and Music

Prior literature suggests that observers may be influenced by a simultaneously presented auditory stimulus when asked to make judgments of a visual stimulus (Gerdes et al., 2014). For example, short excerpts of music have been shown to modulate judgments of the emotion depicted in facial expressions or other images (Logeswaran & Bhattacharya, 2009; Marin et al., 2012). Additionally, Braun Janzen et al. (2022) demonstrated that museum visitors rated a Kandinsky painting as depicting more positive emotional valence if they listened to music they liked rather than disliked while viewing the painting. When rating the emotions depicted in a combined audiovisual film scene, semantically incongruent music decreased the rating of emotions in the visual scene (Rosenfeld & Steffens, 2019). Similarly, prior studies have investigated the influence of auditory stimuli on judgments of properties of visual stimuli: For example, listening to positively valenced music is associated with perceiving a visual stimulus to be brighter (Bhattacharya & Lindsen, 2016). These studies of the depicted emotional valence and visual brightness coincide with prior work indicating that perceptual judgments of visual *features* are also biased by other concurrent stimuli.

In addition to these studies of feature judgments, prior work indicates that listeners report stronger felt emotions when presented with emotionally congruent auditory and visual stimuli together, than when each stimulus is presented in isolation (Baumgartner, Esslen, et al., 2006; Baumgartner, Lutz, et al., 2006; Pan et al., 2019). Similarly, Klein et al. (2021) reported that people like images with low complexity more in the presence of background music, independent of the music's complexity. In sum, this work suggests that, in contrast to prior work presenting two visual stimuli, participants may be unable to faithfully make subjective judgments of a single stimulus in the presence of another when that stimulus is of a different sensory modality.

Influence of Stimulus Duration on Aesthetic Judgments

An additional component of the present work is to test what influence, if any, stimulus duration has on judgments of concurrently presented images and music. While previous research has indicated that observers can reliably report subjectively felt pleasure from an image in as little as 50 ms (Brielmann & Pelli, 2018; Forster et al., 2016; Schwabe et al., 2018), other work has indicated that listeners need at around 500–750 ms to accurately make aesthetic judgments of music (Belfi et al., 2018). Additionally, prior work found that observers cannot faithfully report the single pleasure of a visual stimulus in an array of four stimuli when presented very briefly (200 ms), but that they instead report overall diminished pleasure (Brielmann & Pelli, 2021). This result stands in contrast to similar work that presented multiple images for a longer duration (1.5 s) and found that observers are biased toward the mean pleasure of the entire array (Alwis & Haberman, 2020). When presented with multiple images for a longer period, observers likely had time to attend to each stimulus in a serial manner, which then subsequently resulted in averaging across all stimuli when making a judgment. Here, we expand on this prior work both by introducing two sensory modalities (visual and auditory) as well as two durations (2 s, 5 s).

Theoretical Significance

The present study serves to clarify the nature of when and how sensory inputs from multiple sensory domains are combined in order to make evaluative judgments. Our current study therefore has theoretical implications for how we think about the way sensory information is combined. Brielmann and Pelli (2020, 2021) suggested that rating variance is a product of the rating process and that therefore, pleasures from two simultaneously presented images do not seem to be independently sampled. Here, we present stimuli of two different modalities, making it far more likely that pleasure is being sampled independently from the two sources. One goal of the present work is to provide an additional test for the hypothesis that, when combining pleasures, rating variability systematically decreases during the rating process. The theoretical importance of our study also reaches beyond sensory perception because the existence of separable (hedonic) information from simultaneously presented stimuli forms the basis for making value-based decisions. That is, a better understanding of whether, or under which conditions, observers can make independent judgments of auditory and visual stimuli, will help inform theories of how individuals make value-based decisions about such sensory objects.

At a higher order level, our work also seeks to inform several major theories of aesthetic judgments, particularly on the role played by extra-stimulus or contextual information. One of the major topics of debate in this field surrounds the relative contributions of stimulus features versus external factors to aesthetic judgments of sensory stimuli. Recent theories highlight the important role of context and extra-stimulus features and their influence on aesthetic or hedonic judgments of a stimulus (Brattico & Pearce, 2013; Leder & Pelowski, 2021; Pelowski et al., 2017). For example, prior work has found differences in aesthetic judgments based on the title of an artwork or musical piece (Anglada-Tort et al., 2019; Turpin et al., 2019), the performer of a musical piece (Belfi et al., 2021), whether the composer of a musical piece is familiar or unknown (Fischinger et al., 2018), and biographical information about the composer (Kiernan et al., 2022). This work has indicated that, when asked to make an aesthetic judgment, observers tend to take a holistic approach such that extra-stimulus information does influence their ratings of a stimulus. While theories based on these findings suggest that contextual or external information should contribute to judgments of a single stimulus, here we sought to test whether this biasing by extra-stimulus information is compulsory. That is, when asked to rate a single stimulus in the presence of others (while ignoring the other "distractor" stimuli), can observers faithfully do this? Or, is the contribution of extra-stimulus information obligatory? Thus, the findings from the present work will inform and refine such theories in terms of the role that extra-stimulus information plays in aesthetic judgments.

Whether observers are biased by extra-stimulus information might also be affected by the perceived source of that information. One related concept in the field of multisensory integration is the "unity assumption," which occurs when observers perceive individual unisensory objects as belonging to the same source (for review, see Chen & Spence, 2017). If an individual assumes that two stimuli result from the same source, they are more likely to integrate and perceptually bind crossmodal stimuli. This has been shown in the realm of audiovisual speech, such that individuals show more crossmodal binding for auditory and visual speech if it is perceived as coming from the same speaker. That is, the unity assumption is more likely to occur when two sensory stimuli are perceived as "congruent," for example, a male voice with a male face when perceiving speech sounds (Vatakis & Spence, 2007), or a sound and video of the same musical instrument (Chuen & Schutz, 2016). Even abstract congruencies can influence crossmodal binding, such as auditory pitch and visual size (Parise & Spence, 2009) or timbre and brightness (Wallmark et al., 2021).

When considering how contextual information influences aesthetic judgments, many types of extra-stimulus information are likely to be assumed to come from the same source (e.g., the title of an artwork is "part" of the work itself, or the composer or musical artist created the work). Thus, in prior work which found an influence of contextual information on aesthetic judgments, this biasing by extra-stimulus information could be because this additional information was assumed to be *a part* of the stimulus being evaluated. However, the present study tests extra-stimulus information that is perhaps unlikely to induce the unity assumption. Thus, the results of the present work will inform both theories of aesthetic judgments (i.e., in what conditions aesthetic judgments are influenced by extra-stimulus information) as well as multisensory integration.

The Present Study

Therefore, in the present work, we sought to test whether observers can independently judge the pleasure of a single stimulus (i.e., a single image or a single musical excerpt) when concurrently presented with a stimulus of the other sensory modality. Additionally, we sought to test whether observers can accurately average their pleasure responses to two stimuli when asked to rate the combined pleasure of the image and music. Here, participants viewed images of paintings and heard excerpts of instrumental music and were asked to rate the pleasure they felt either from a single stimulus or the combined pleasure of both stimuli. Additionally, we sought to investigate whether stimulus duration has an influence, and tested this question by repeating the experiment with short (2 s) and longer (5 s) stimulus durations.

Based on prior literature (Brielmann & Pelli, 2020), we tested multiple models for how observers might rate either the single or combined pleasure of images: When asked to report the pleasure of a single stimulus, we evaluated four possible models: observers might (a) faithfully report the pleasure (*faithful model*), (b) average the pleasure of two stimuli (*compulsory averaging model*), (c) partially bias their pleasure report toward the irrelevant stimulus (*linear model*), or (d) partially bias their pleasure report based on the stimulus modality (music or image; *linear modality model*). Previous research testing the independence of judging two images (Brielmann & Pelli, 2020, 2021) did not include this fourth model, which accounts for differences in stimulus modality (because they only studied visual stimuli). Previous work has indicated that, when asked to report stimulus features of multiple simultaneously presented stimuli, participants tend to report the average (Brady & Alvarez, 2011; Haberman & Whitney, 2009; Maule et al., 2014). In contrast, when asked to report the *pleasure* of two images, observers are able to faithfully report the pleasure (Brielmann & Pelli, 2020). Therefore, given that the present work is investigating pleasure ratings, we predicted that our data would be more likely to be fit by a faithful model than a compulsory averaging model. However, a key difference in the present study is that the present work includes two sensory domains—auditory and visual. Therefore, we have added the fourth model that assigns weight based on the modality. If stimulus modality is particularly important, then, we would expect this final model to be a better fit for our data than the faithful model.

Similarly, and in line with previous work (Brielmann & Pelli, 2020), we tested four models that predict observers' responses when asked to report the combined pleasure of two stimuli: (a) faithfully averaging the pleasure of both stimuli (*faithful model*), (b) reducing the weight of stimuli with extreme pleasure values (*compressive model*), (c) making an average pleasure rating that is lower at the low end and higher at the high end than predicted based on ratings for the stimuli presented in isolation (*expansive *model*), or (d) make ratings that are differently weighted based on stimulus modality (*linear modality model*). Our predictions again are similar to those described above—based on prior studies we predict that the faithful model will be the best fit for our data, unless, that is, stimulus modality plays an important role (in which case we predict the linear modality model would be the best fit).

Method

Participants

Participants were undergraduate students at Missouri University of Science and Technology who completed the experiments as part of their course credit. We were unable to conduct a power analysis to determine sample size because we do not employ traditional frequentist analyses and use specially designed models to fit our data. Since these models additionally differ in their complexity, i.e., number of free parameters, a simple frequentist test for differences between errors would be misleading, too. However, our current work is modeled very closely to our prior work (Brielmann & Pelli, 2020, 2021). We therefore set our target N at 30 participants, based on the sample sizes that have been used successfully in the past on a very similar task. The similar sample size additionally allows us to make direct comparisons between our and these previous findings. Note that we also employ our main analyses on the level of individual participants and that we can show that our results hold for the majority of individual participants, too. We additionally ensure the reliability of our results by replicating our experiment twice. To account for participant attrition due to potentially inattentive participants, we sought to recruit 40–50 participants per study.

Experiment 1

Of the 42 participants who completed Experiment 1, five were excluded from the analyses for failing the attention check. This left a total of 37 participants who completed the experiment (26 identified as male, 10 as female, and one as "other"). Their mean age was 19.8 years ($SD = 2.1$), they had an average of 13.66 years ($SD = 1.3$) of education, 4.5 years ($SD = 3.42$) of formal musical training, and 1.3 years ($SD = 1.71$) of formal art training.

Experiment 2

Of the 48 participants who completed Experiment 2, nine were excluded from the analyses for failing the attention check. This left a total of 37 remaining participants, (27 identified as male and 12 as female). Their mean age was 19.4 years ($SD = 1.3$), they had on average 13.7 years of education ($SD = 1.3$), 3.7 years ($SD = 3.0$) of formal musical training, and 2.0 years ($SD = 2.8$) of formal art training. We repeated all analyses for both experiments with the removed subjects included, and the results remained unchanged. Therefore, below we report the results without these participants.

Stimuli

Stimuli consisted of 18 musical excerpts and 18 images of paintings. All stimuli were selected from our prior work (Belfi, 2019; Belfi et al., 2018, 2019). Musical stimuli consisted of instrumental musical excerpts selected from two genres: "classical" and "jazz." Classical excerpts were chosen from 19th-century small ensemble music of the Romantic era. Jazz music consisted of 1960s music with bop or post-bop elements. Ensembles and instrumentation were all traditional jazz, consisting of piano, guitar, saxophone, brass, and so forth. No particularly well-known pieces were selected, in an effort to ensure that all pieces were unfamiliar. Musical pieces had previously been rated on their aesthetic appeal (Belfi et al., 2018). In order to span the range of aesthetic appeal in the present study, we selected the nine musical pieces that were the most highly rated (e.g., most-liked) and the nine musical pieces that were the least highly rated (e.g., least-liked).

Visual stimuli consisted of images of paintings that had been used in our prior work (Belfi et al., 2019). Images were selected from the Catalog of Art Museum Images Online database (http://www.oclc.org/camio). These images are high-quality photographs of paintings from a variety of cultural traditions (American, Asian, and European) and time periods (15th century to the present). Commonly reproduced images and particularly well-known paintings were not included in order to minimize familiarity. As with musical stimuli, we selected the top nine most-liked and least-liked images based on previous ratings.

Procedure

All procedures were conducted in compliance with the American Psychological Association Ethical Principles and were approved by the Institutional Review Board of the University of Missouri. This study was conducted online using Gorilla Online Experiment Builder (http://www.gorilla.sc) to create and host the experiments described here (Anwyl-Irvine et al., 2020). Data were collected throughout the year 2021. Procedures generally followed those from a prior similar experiment using two images (Brielmann & Pelli, 2020) with some modifications for the present experiments. There were two blocks in the main task: a *precued* block and a *postcued* block. In the precued block, a cue was presented before the stimuli that indicated whether to rate the pleasure of the music, the image, or the combined pleasure of both stimuli. The cue was a word presented in the middle of the screen that said "music," "image," or "both." In the postcued block, the same cues were presented but *after* the stimuli. The order of the precued and postcued blocks was counterbalanced across participants. At the end of these two blocks, participants completed a final baseline rating task. In this baseline task, only one stimulus appeared at a time and participants rated each stimulus in isolation. Baseline ratings were blocked, such that participants rated all images first and then all music (or vice versa; order of image and music baseline ratings was counterbalanced across participants).

First, participants were instructed on the task (see the online supplemental materials for complete instructions). Next, they completed six practice trials, one for each possible cue (music, image, both), once precued and once postcued. Next, participants proceeded to the main experiment. Participants completed one precued block and one postcued block. In each block, each stimulus was presented once as *target* (i.e., the stimulus that is cued), once as *distractor* (i.e., the non-cued stimulus), and once as part of a pair whose combined pleasure was rated (total $N = 54$ trials per block). At the beginning of each block, participants were told whether the cue would come before or after the stimuli. Finally, participants completed the baseline rating blocks. To rule out the possibility that baseline ratings at the end of the experiment were systematically corrupted by such sequence effects, we assessed whether the ratings of a participant or for a particular stimulus changed during the time course of the experiment (see the online supplemental materials for details). Median correlations between trial number and ratings across participants or stimuli never exceeded an absolute value of 0.12 in any of the experiments, confirming that no overall systematic sequence effect could bias the relationship between ratings in the main experiment and baseline ratings.

Trial timelines for the different types of blocks are illustrated in Figure 1. Stimuli were presented for 2,000 ms (Experiment 1) or 5,000 ms (Experiment 2). Images were presented in the center of a white screen. After the stimuli, participants were instructed to "Rate how much pleasure you felt from the stimulus, where 1 is *no pleasure at all* and 9 is *very intense pleasure*." Ratings were made by using a mouse click to select one of nine buttons labeled 1–9 on the screen. One attention check question was included, which was an audio stimulus that instructed participants to select a specific number on the rating scale.

At the end of the experiment, participants also completed two questionnaires to assess their overall pleasure responses to both music and visual images: the Barcelona Music Reward Questionnaire (BMRQ; Mas-Herrero et al., 2013) was used to assess musical reward, whereas the Aesthetic Responsiveness Assessment (AReA; Schlotz et al., 2020) was used to assess responsiveness to visual art. Finally, participants completed a brief demographics questionnaire and were asked about their age, gender, years of education, years of formal musical training, and years of formal visual arts training.

Analysis

We refer to trials in the main experiment where participants rated the pleasure of one target stimulus in the presence of a second distractor stimulus (e.g., rating the music while an image was present) as *one-pleasure* trials. Trials in which participants rated the combined pleasure of both stimuli are referred to as *combined-pleasure* trials. Finally, the baseline trials at the end of the experiment are referred to as *single-pleasure* trials.

Our main analyses replicate the ones applied to data for two simultaneously presented images (Brielmann & Pelli, 2020). That is, we use leave-one-out-cross-validation (LOOCV) with the root mean square errors (RMSEs) as our measure of goodness of fit to evaluate all models on an individual participant basis. As in the previous report, we model our data as a linear transformation of the weighted sum of both stimuli's single pleasures:

$$\hat{P} = a + b(wP_1 + (1 - w)P_2) \quad (1)$$

where $\hat{P}$ is the estimated rated pleasure, w is the target weight, $0.5 \leq w \leq 1$, and a and b are constants. For one-pleasure trials, P_1 represents the target's and P_2 the distractor's single pleasure; for combined-pleasure trials, P_1 represents the image's single pleasure and P_2 the music's single pleasure.

In addition to the above models, we here also consider a modality-specific weighted sum of the stimuli's single pleasures:

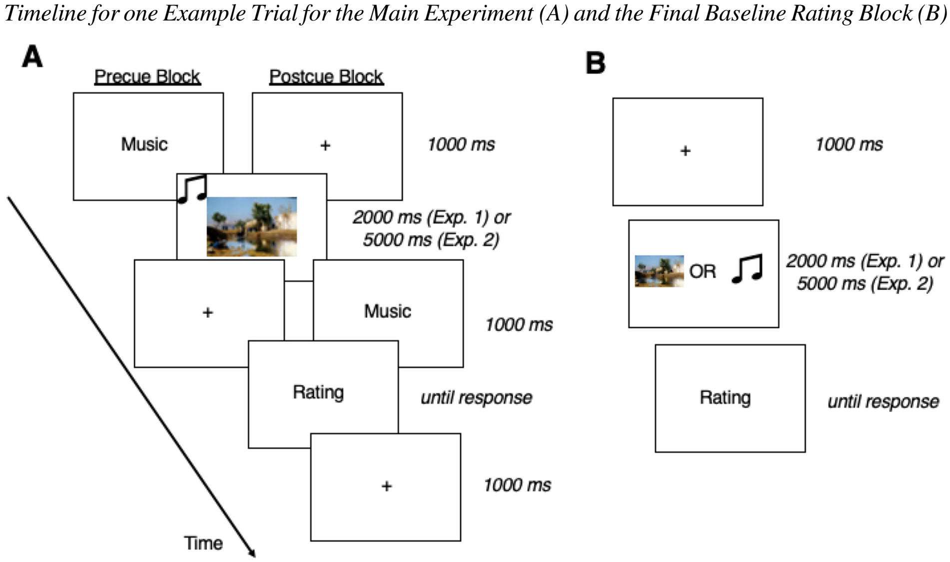
$$\hat{P} = wP_i + (1 - w)P_{\mathrm{m}} \quad (2)$$

where P_i represents the image's single pleasure and P_{m} the music's single pleasure. Note that this equation is equivalent to Equation 1 in combined-pleasure trials.

We fit four models to each participant's data. Three were based on Equation 1: (a) the faithful model ($w = 1$; $a = 0$; $b = 1$), (b) the compulsory averaging model ($w = 0.5$; $a = 0$; $b = 1$), and (c) a linear model based on target assignment ($a = 0$; $b = 1$; where w is a free parameter). The fourth model (see Equation 2) is also a linear model but assigns the weight w based on the stimulus modality (image vs. music).

For combined-pleasure trials, we considered three models that are variations of Equation 1 assuming equal weighting of both stimulus pleasures ($w = 0.5$): (a) the faithful model with $a = 0$; $b = 1$; (b) the compressive model with $0 < b < 1$, $a > 0$; (c) the expansive model

Figure 1
Timeline for one Example Trial for the Main Experiment (A) and the Final Baseline Rating Block (B)



Note. Image from "Fellah Women Drawing Water," by Jean-Leon Gerome, c. 1873–75, Clark Art Institute (www.clarkart.edu). In the public domain. See the online article for the color version of this figure.

with $b > 1$; $a < 0$. In contrast, w is a free parameter for the fourth model we consider, the modality-specific linear model that weighs single pleasures based on stimulus modality (Equation 2).

Like previous studies (Brielmann & Pelli, 2020; Haberman et al., 2015) we report Cronbach's α for the average absolute error per participant as measure of reliability. Cronbach's α measures intercorrelations among items to gauge internal consistency. That is, for one-pleasure trials, we calculated the absolute deviation of each trial's pleasure rating from the single-pleasure rating (i.e, baseline rating) of the to-be-rated stimulus. For combined-pleasure trials, we calculated absolute deviation of pleasure reports from the mean across single pleasures of the two presented stimuli. We then averaged absolute errors per participant. Then, Cronbach's α values were used to compute an upper bound for the correlation between errors in the two kinds of trials, using Equation 3. In this equation, X and Y are two random variables (in our case, the errors in two kinds of trials) and α is Cronbach's α (Nunnally, 1970):

$$r_{XY} \geq \sqrt{\alpha_X \alpha_Y} \tag{3}$$

We use this upper bound to evaluate the correlations of the errors of trials with the same instructions (rate one or rate combined) by correlating pre- versus postcued errors and, inversely, the errors of trials with the same cue timing (pre or postcued) by correlating one versus combined-pleasure errors. We again use the average error per stimulus and participant as the basis for these correlations.

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in this study. Raw data, analysis files, and details of all packages used are available at https://github.com/aenneb/image_and_music_pleasure. Materials are copyrighted and therefore available upon request. We analyzed the data with Python 3.8.1 running in Spyder 4.1.4. This study's design and its analysis were not preregistered.

Results

Experiment 1: Short Stimulus Presentation

Reliability and Correlations Between Different Trial Types

Cronbach's α varied considerably between participants but was, on average, high across trial types (see Figure 2A). In contrast to what has been reported for two images, reliability tended to be higher for combined-pleasure trials in our mixed-modality experiment (precued $\alpha = 0.81$, 95% CI [0.72, 0.89]; postcued $\alpha = 0.84$, [0.76, 0.91]) than for one-pleasure trials (precued $\alpha = 0.45$ [0.17, 0.68]; postcued $\alpha = 0.66$[0.49, 0.80]). The non-overlapping confidence intervals (CIs) for precued trials indicate that this difference is meaningful at least in this case, when extending the standard benchmark of a less than 5% chance of overlap in estimated statistics as meaningful difference to the case of Cronbach's α. We also calculated Cronbach's α separately for trials with music versus image as a target (separately for pre- and postcued trials) but did not find a general pattern of difference between music and image trials (see the online supplemental materials for details).

The fact that reliability of combined ratings was higher than for one-target ratings is expected under the assumption that these ratings represent an averaged report of two individually sampled pleasure values. That is, the averaging of two independent samples reduces the error because the random noise associated with each sample is compensated when averaging both samples. Random noise will, across several samples, deviate evenly in either direction from the ground truth. Note that this result stands in contrast to equal (if not higher) reliability of single-pleasure reports for two images (Brielmann & Pelli, 2020).

Inserting the above Cronbach's α values into Equation 3, the maximal expected correlations are 0.55 between pre- and postcued one-pleasure trials, 0.83 between pre- and postcued combined-pleasure trials, 0.61 between one- and combined-pleasure precued trials, and 0.75 between one- and combined-pleasure postcued trials. Figure 2B illustrates the correlations of errors between trial types. They were well below the maximally achievable correlations and were considerably lower in our study than previously reported for two images. Nonetheless, the relative differences between trial types were similar in both studies. Error correlations were higher between trials with different cue timings (i.e., correlations between pre- vs. postcued trials) that instructed the participant to rate one pleasure ($r = .47$, 95% CI [0.43, 0.51]) or the combined pleasure ($r = .12$, [0.07, 0.18]) than between trials with the same cue timing and different instructions (i.e., correlations between one-pleasure and combined-pleasure trials for precued $r = .01$, [−0.04, 0.07] and postcued $r = .03$, [−0.03, 0.08]). These analyses collapsed across trials with music as target and images as target. Therefore, we additionally calculated error correlations between pre- and postcued trials for trials with images as target ($r = .39$, [0.32, 0.45]) and trials with music as target ($r = .53$, [0.47, 0.58]) separately. These correlations were within the same range as the correlations for trials with the same instructions.

People Can Report the Pleasure of Simultaneously Presented Images and Music Independently

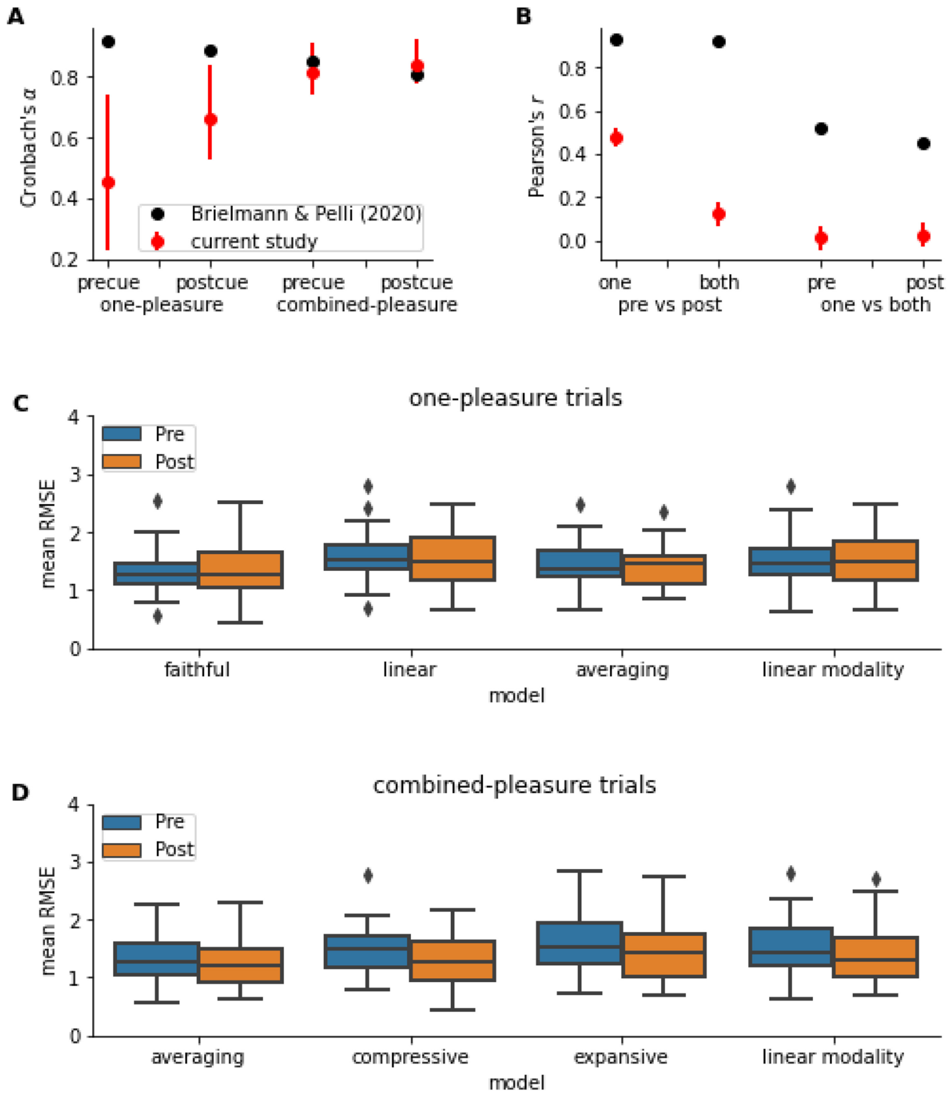
Figure 2C compares the outcomes of all models for one-pleasure trials. On average, the faithful model fit the data in pre- (mean RMSE = 1.31) and postcued trials (1.32) best and it was the nominally best-fitting model according to RMSE for the majority of participants in pre- (26/37) and postcued trials (24/37), too. Note that even though average RMSEs are similar for all models, the faithful model is the one to be favored because it has fewer (i.e., 0) free parameters than any alternative model. Thus, people report the pleasure of one image or musical piece in an unbiased fashion even when accompanied by music or an image, respectively.

People Can Report the Average Pleasure Across an Image and Music

Figure 2D compares the outcomes of LOOCV for combined-pleasure trials for all models. On average, the faithful averaging model fit the data in pre- (mean RMSE = 1.22) and postcued trials (1.30) best and it was the best-fitting model according to RMSE for the majority of participants in pre- (26/37) and postcued trials (21/37), too. We again favor the simplest, parameter-free model, here, because it achieves similar average and even nominally lower individual error values than alternative, more complex models. No other model fit a similar number of participants best (see the online supplemental materials for detailed counts). Thus, people's judgments of the combined pleasure they experience based on an image accompanied by music are best captured as the arithmetic average of the music's and the image's single-pleasure ratings.

Figure 2

Reliability, Correlations, and Model Fitting Results for Experiment 1



Note. A: Cronbach's α for rating errors as a measure of reliability per trial type. B: Pearson's *r* as measure of correlation between rating errors between trial types. A, B: Red dots represent data of the current study, black ones of a previous study on two simultaneously presented images (Brielmann & Pelli, 2020). Error bars indicate 95% confidence intervals. C, D: Boxplots of root mean square errors (RMSEs) for each model for precued trials (left; blue) and postcued trials (right; orange). Boxplot properties are the default setting of the python package *seaborn*. C: RMSEs for one-pleasure trials. D: RMSEs for combined-pleasure trials. See the online article for the color version of this figure.

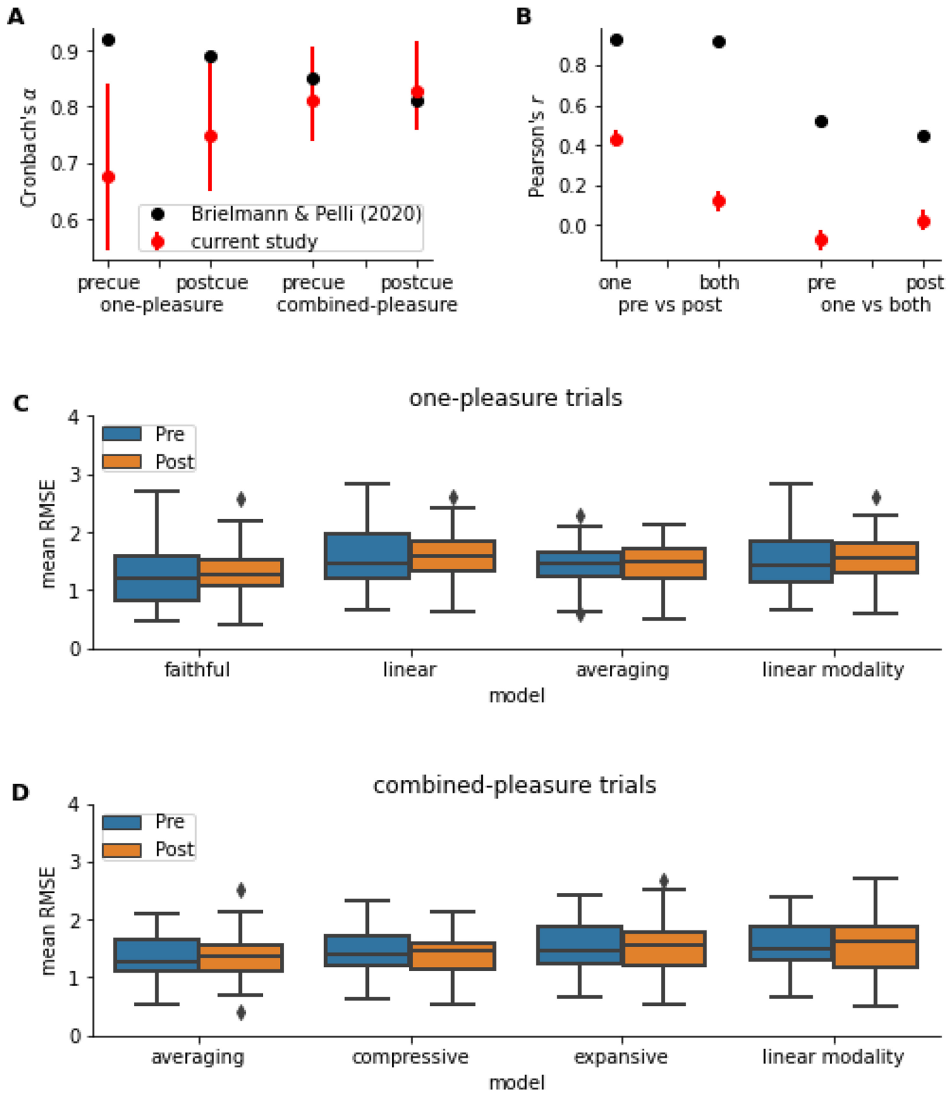
We also investigated whether there were systematic relationships between the goodness of model fits and/or parameter estimates and participants' BMRQ and AReA scores. By and large, there were no meaningful and consistent correlations. We report the detailed analyses and results in the online supplemental materials.

Experiment 2: Long Stimulus Presentation

The analyses for Experiment 2 were identical to those for Experiment 1. As illustrated in Figure 3, the increased stimulus duration did not change the results. Cronbach's α was again slightly lower in one-pleasure trials (precued $\alpha = 0.68$, 95% CI [0.51, 0.81]; postcued $\alpha = 0.75$, [0.62, 0.85]) than in combined-pleasure trials (precued $\alpha = 0.81$, [0.71, 0.89]; postcued $\alpha = 0.83$, [0.74, 0.90]). However, the overlapping confidence intervals suggest that the differences in this experiment were less robust, even though the overlap in CIs does not allow us to draw definitive conclusions about the presence or absence of meaningful differences. We again also calculated Cronbach's α separately for the different target modalities (image and music) with similar results to the combined analyses (see the online supplemental materials). Error correlations were again higher between trials with different cue timings (i.e., correlations between

Figure 3

Reliability, Correlations, and Model Fitting Results for Experiment 2



Note. A: Cronbach's α for rating errors as a measure of reliability per trial type. B: Pearson's *r* as measure of correlation between rating errors between trial types. A, B: Red dots represent data of the current study, black ones of a previous study on two simultaneously presented images (Brielmann & Pelli, 2020). Error bars indicate 95% confidence intervals. C, D: Boxplots of root mean square errors (RMSEs) for each model for precued trials (left; blue) and postcued trials (right; orange). Boxplot properties are the default setting of the python package *seaborn*. C: RMSEs for one-pleasure trials. D: RMSEs for combined-pleasure trials. See the online article for the color version of this figure.

pre- vs. postcued trials) that instructed the participant to rate one pleasure ($r = .43$, [0.39, 0.47]) or the combined pleasure ($r = .12$, [0.07, 0.17]) than between trials with the same cue timing and different instructions (i.e., correlations between one-pleasure and combined-pleasure trials for precued $r = -.07$, [−0.12, −0.02] and postcued $r = .03$, [−0.03, 0.08]). Additionally, we calculated error correlations between pre- and postcued trials for trials with images as target ($r = .44$, [0.38, 0.50]) and trials with music as target ($r = .43$, [0.36, 0.48]) separately. These correlations were within the same range as the correlations for trials with the same instructions.

The faithful model achieved the on average lowest error in single-pleasure trials (precued RMSE = 1.26; postcued RMSE = 1.32) and was the best-fitting model for most participants (29/39 precued; 27/39 postcued). Also, the faithful averaging model had the lowest average RMSE across participants in combined-pleasure trials (both RMSEs = 1.37) as well as being the best-fitting model for a majority of individual participants (24/39 precued; 22/39 postcued). As before, this conclusion is underscored by the fact that the best-performing model is the one without free parameters while all alternative models contain at least one free parameter.

Discussion

In the present work, we sought to investigate whether individuals can report the pleasure of a single item (i.e., a musical excerpt or image of a painting) in the presence of a stimulus in the other modality—That is, can people accurately judge the pleasure of an image while listening to music, and vice versa? Our results indicate that people *can* do this: Participants reported the pleasure of an image or a piece of music faithfully, even in the presence of the other. Additionally, we sought to test whether people can accurately average the pleasure of two stimuli when asked to do so. Again, our results indicated that people can faithfully report the average: When prompted to rate their combined pleasure, people's reports are best described as the arithmetic mean of the image's and the music's individually rated pleasures. This pattern of results was evident whether people knew what to rate before stimulus onset (i.e., the precued trials) or after stimulus offset (i.e., the postcued trials) as well as with shorter (2 s) and longer (5 s) exposure durations. To further illustrate the robustness of this effect, we also replicated these main results (using the shorter, 2 s stimulus presentation) in an online sample of participants (see Experiment 3 in the online supplemental materials).

These findings also replicate those previously reported for studies with two simultaneously presented images (Brielmann & Pelli, 2020). This suggests that people can faithfully encode and report the pleasures of at least two objects, independent of correspondence in stimulus modality. While our main results replicated those of image–image pairings, we did find some differences between the current and the previous image-only studies. For one, rating reliability was higher for combined-pleasure ratings (as compared to one-pleasure ratings) in the studies reported here, as one would expect when pleasures are averaged across two independent observations, whereas it was no different for one- versus combined-pleasure ratings of two images. This suggests that late-stage response noise arises when averaging pleasures of the same modality but not when combining pleasures of two different modalities. In addition, correlations between average errors in different trial types were considerably lower in our study than they could have been based on their reliability. This, too, stands in contrast to previous findings with two images where correlations were at or close to ceiling. Hence, rating behavior varied more as a function of cue timing and requested report (one vs. combined pleasure) when people experienced an image and music rather than two images. This suggests that participants used different strategies (implicitly or explicitly) in the different trial types. Nonetheless, our modeling results show that all of these presumably different strategies resulted in faithful pleasure reports of the target stimulus and the average across stimuli in one- and combined-pleasure trials, respectively.

At first glance, it might seem that this work stands in some contrast to prior work investigating how the presence of music influences judgments of visual stimuli. That is, prior work found that simultaneously presented music does have an influence on judgments of images. For example, hearing a "happy" song while viewing a "happy" face was associated with increased judgments of positive valence, as compared to hearing the music alone (Pan et al., 2019) or seeing the face alone (Baumgartner, Lutz, et al., 2006). However, our study was distinct from these prior works in the type of response participants made. Here, we were interested in the amount of pleasure felt in response to the stimuli, rather than judgments of the emotional *content* of the stimuli. This contrast between studies of pleasure ratings versus ratings of stimulus features is similar to that found in Brielmann and Pelli's findings (2020). While they found that observers could reliably report the *pleasure* of a single item in the presence of others, this stands in contrast to prior work demonstrating that ratings of a single item's *perceptual properties* are biased toward the average of the set, rather than faithfully reported (Brady & Alvarez, 2011; Haberman & Whitney, 2009; Maule et al., 2014). Another difference is that here, we randomly combined pairs of most-pleasing and least-pleasing music and images (as determined by ratings from prior work), whereas prior work chose images and music to depict certain emotions or stimulus characteristics (e.g., happiness vs. sadness, or complex vs. simple). Thus, it seems to be that rating features of the stimulus (whether perceptual or emotional) may be more at risk to bias from other simultaneously presented stimuli than rating one's felt pleasure in response to a stimulus. An additional difference between the present study and prior, similar work, is that previous studies tend to focus on the influence of "background" music (Braun Janzen et al., 2022; Klein et al., 2021). These prior studies presented a single extended musical excerpt, for minutes or longer, while the visual stimuli changed on screen (Baumgartner, Esslen, et al., 2006; Baumgartner, Lutz, et al., 2006). While we found no difference between 2 s and 5 s stimulus presentation, such very extended presentation durations may lead to one stimulus influencing judgments of the other. Therefore, it could be the case that one's rating of an image would be influenced by a concurrent musical stimulus if presented for an extended period of time.

At first glance, our work also stands in contrast with other prior work which has found that context and extra-stimulus information does influence an observer's aesthetic judgment of an object. For example, previous research has indicated that information about a musical performer (Shank et al., 2022) or title of an artistic work (Millis, 2001; Turpin et al., 2019) can have an influence on one's aesthetic judgment of that work. However, in these cases, this extra-stimulus information can be presumed to be a part of the stimulus itself and to have originated from the same source (i.e., the artist). The present findings, therefore, have important implications toward both theories of aesthetic judgments as well as multisensory integration more broadly—here, we find that observers are able to make independent judgments of multisensory stimuli, when those two stimuli are not likely to induce the unity assumption (i.e., the assumption that two individual unisensory stimuli are derived from the same source). An interesting direction for future research would be to further probe this effect by presenting musical and visual stimuli that are purported to originate from the same source (e.g., a multisensory art installation created by a single artist) to see whether inducing the unity assumption affects an observers' ability to independently make hedonic judgments of a stimulus.

This point also relates to the question of how stimulus features may influence the effect seen here. That is, specific features of the stimuli might induce certain conditions that affect whether an observer is likely to assess a stimulus independent of others in the environment. The stimuli in the present study were chosen to be those that were rated as having high or low aesthetic appeal, in order to get a broad range of liked versus disliked stimuli. While we attempted to select stimuli that spanned the extremes of this range, as with many studies of aesthetic appeal, ratings tended to be more moderate for our stimuli (for a full list of stimuli and normative ratings, see the online supplemental materials). Nevertheless, it is possible that the present results were driven by the particular stimuli chosen here. If we were to have chosen more "neutral" images or musical clips, ratings may have been more likely to be influenced by the other modality. One interesting point to note regarding this is that the musical stimuli had a smaller range of normative ratings than visual stimuli. That is, musical stimuli were more "neutral" than visual stimuli, overall. Despite this, we did not see any differences in the ability for music to influence judgments of images or vice versa (which would be seen in the fourth model we tested, which accounts for modality). Therefore, it seems to be that neutral stimuli are not any more likely to be biased by more "extreme" stimuli.

Another factor that might contribute to whether or not stimuli are judged independently is semantic congruence. An emerging area in the field is the study of narratives or other semantic content evoked by music. Recent work indicates that listeners reliably imagine consistent narratives while listening to musical pieces (Margulis, 2017; Margulis et al., 2022; McAuley et al., 2021). It might be the case that, if musical and visual stimuli share such semantic content, ratings of one modality may be influenced by a stimulus in the other. For example, if a musical excerpt is likely to evoke imagery of a battle, and an observer is viewing a painting of a battle, their ratings of the painting may be more likely to be influenced by the music than if they were listening to an excerpt that evoked, for example, a sunrise. In addition to narratives, music evokes other crossmodal correspondences. For example, listeners consistently match music to tastes based on certain musical features: Lower-pitched and legato notes sound "bitter," while staccato notes sound "salty" (Guetta & Loui, 2017). Other work has identified consistent patterns of associations between music and color: For example, faster pieces in a major mode tend to be associated with brighter colors (e.g., yellow) while slower pieces in a minor mode are associated with darker colors (e.g., blue; Palmer et al., 2013; Whiteford et al., 2018). It is likely the case that the stimuli used in the present study did not display such crossmodal correspondences and therefore were less likely to influence ratings of the other stimulus modality.

This work also has important theoretical and practical implications for applied fields. Perhaps an obvious example is applications to the advertising industry. While traditionally this industry has relied on visual marketing techniques, there has been an increasing interest in audio branding in recent years (Minsky & Fahey, 2017). Marketers have suggested that audio information (e.g., music or other sounds) can influence the likelihood for customers to purchase or be interested in a product. For example, recent work has shown that individuals are more likely to buy an item if it is advertised with familiar, versus unfamiliar, music (Anglada-Tort, Schofield, et al., 2022). Similarly, a recent theoretical framework, The Behavioral Economics of Music, has proposed that decision-making about how much a listener enjoys a piece of music (and other aesthetic judgments) should be considered similar to other sorts of economic decisions (Anglada-Tort, Masters, et al., 2022). This framework considers contextual effects on musical judgments (such as influences of the composer, title, or performer of a piece) as a type of heuristic, similar to the affect heuristic—that is, information about the composer or performer evokes an emotion and that in turn is what influences one's judgments about the piece. Our current work suggests that, perhaps, observers are less likely to engage heuristic processes when asked to make evaluations of two stimuli in different modalities, or that they are able to ignore the affect response evoked by the irrelevant stimulus.

One important caveat of the present work is that it can only be generalized to certain populations and certain types of stimuli. To generalize beyond an undergraduate population, we replicated Experiment 1 with participants recruited online via Prolific (see the online supplemental materials). Prior work has indicated that online samples tend to be more diverse than undergraduate student samples, and has additionally shown that Prolific samples tend to be more diverse than samples recruited on other online platforms such as Amazon's Mechanical Turk (Peer et al., 2017). However, we limited this online sample to U.S.-based participants, to match our U.S.-based undergraduates. Therefore, we cannot generalize our results beyond a U.S.-based population. We here aim to make claims about this population's ability to report pleasure from simultaneously presented visual and auditory stimuli and therefore did not recruit additional expert populations, e.g., musicians or art historians, and leave it up to future studies to examine whether our results will generalize to such special populations. Additionally, it is important to note that for our musical stimuli, we used classical and jazz musical excerpts that come from the Western musical tradition. This is a limitation that does not allow generalizability beyond Western listeners of Western music (Baker et al., 2020).

To conclude, our results show that individuals can give unbiased reports of their felt pleasure from an image or a piece of music, even when in the presence of another stimulus. Additionally, they are also able to faithfully report the average pleasure of both stimuli when asked to do so. These results replicate findings with two image pairings and extend this work into judgments of multisensory stimuli. Future research could further extend these findings into other sensory domains (e.g., touch, smell) and continue to probe the question of stimulus duration in order to test the limits of one's ability to independently track the pleasure of simultaneous stimuli.

References

Alwis, Y., & Haberman, J. M. (2020). Emotional judgments of scenes are influenced by unintentional averaging. *Cognitive Research: Principles and Implications*, *5*(1), Article 28. https://doi.org/10.1186/s41235-020-00228-3

Anglada-Tort, M., Masters, N., Steffens, J., North, A., & Müllensiefen, D. (2022). The behavioural economics of music: Systematic review and future directions. *Quarterly Journal of Experimental Psychology*. Article 17470218221113760. https://doi.org/10.1177/17470218221113761

Anglada-Tort, M., Schofield, K., Trahan, T., & Müllensiefen, D. (2022). I've heard that brand before: The role of music recognition on consumer choice. *International Journal of Advertising*, *41*(8), 1567–1587. https://doi.org/10.1080/02650487.2022.2060568

Anglada-Tort, M., Steffens, J., & Müllensiefen, D. (2019). Names and titles matter: The impact of linguistic fluency and the affect heuristic on aesthetic and value judgements of music. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(3), 277–292. https://doi.org/10.1037/aca0000172

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baker, D. J., Belfi, A., Creel, S., Grahn, J., Hannon, E., Loui, P., Margulis, E. H., Schachner, A., Schutz, M., Shanahan, D., & Vuvan, D. T. (2020). Embracing anti-racist practices in the music perception and cognition community. *Music Perception*, *38*(2), 103–105. https://doi.org/10.1525/mp.2020.38.2.103

Baumgartner, T., Esslen, M., & Jäncke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, *60*(1), 34–43. https://doi.org/10.1016/j.ijpsycho.2005.04.007

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures. *Brain Research*, *1075*(1), 151–164. https://doi.org/10.1016/j.brainres.2005.12.065

Belfi, A. M. (2019). Emotional valence and vividness of imagery predict aesthetic appeal in music. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 128–135. https://doi.org/10.1037/pmu0000232

Belfi, A. M., Kasdan, A., Rowland, J., Vessel, E. A., Starr, G. G., & Poeppel, D. (2018). Rapid timing of musical aesthetic judgments. *Journal of Experimental Psychology: General*, *147*(10), 1531–1543. https://doi.org/10.1037/xge0000474

Belfi, A. M., Samson, D. W., Crane, J., & Schmidt, N. L. (2021). Aesthetic judgments of live and recorded music: Effects of congruence between musical artist and piece. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.618025

Belfi, A. M., Vessel, E. A., Brielmann, A., Isik, A. I., Chatterjee, A., Leder, H., Pelli, D. G., & Starr, G. G. (2019). Dynamics of aesthetic experience are reflected in the default-mode network. *NeuroImage*, *188*, 584–597. https://doi.org/10.1016/j.neuroimage.2018.12.017

Bhattacharya, J., & Lindsen, J. P. (2016). Music for a brighter world: Brightness judgment bias by musical emotion. *PLoS One*, *11*(2), e0148959. https://doi.org/10.1371/journal.pone.0148959

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392. https://doi.org/10.1177/0956797610397956

Brattico, E., & Pearce, M. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 48–61. https://doi.org/10.1037/a0031624

Braun Janzen, T., de Oliveira, B., Ventorim Ferreira, G., Sato, J. R., Feitosa-Santana, C., & & Vanzella, P. (2022). The effect of background music on the aesthetic experience of a visual artwork in a naturalistic environment. *Psychology of Music*, *51*(1), 16–32. https://doi.org/10.1177/03057356221079866

Brielmann, A. A., & Pelli, D. G. (2018). Aesthetics. *Current Biology*, *28*(16), R859–R863. https://doi.org/10.1016/j.cub.2018.06.004

Brielmann, A. A., & Pelli, D. G. (2020). Tracking two pleasures. *Psychonomic Bulletin & Review*, *27*(2), 330–340. https://doi.org/10.3758/s13423-019-01695-6

Brielmann, A. A., & Pelli, D. G. (2021). The pleasure of multiple images. *Attention, Perception, & Psychophysics*, *83*(3), 1179–1188. https://doi.org/10.3758/s13414-020-02175-z

Chen, Y.-C., & Spence, C. (2017). Assessing the role of the ‘unity assumption’ on multisensory integration: A review. *Frontiers in Psychology*, *8*, Article 445. https://doi.org/10.3389/fpsyg.2017.00445

Chuen, L., & Schutz, M. (2016). The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude envelope cues. *Attention, Perception, & Psychophysics*, *78*(5), 1512–1528. https://doi.org/10.3758/s13414-016-1088-5

Fischinger, T., Kaufmann, M., & Schlotz, W. (2018). If it’s Mozart, it must be good? The influence of textual information and age on musical appreciation. *Psychology of Music*, *48*(4), 579–597. https://doi.org/10.1177/0305735618812216

Forster, M., Leder, H., & Ansorge, U. (2016). Exploring the subjective feeling of fluency. *Experimental Psychology*, *63*(1), 45–58. https://doi.org/10.1027/1618-3169/a000311

Gerdes, A. B. M., Wieser, M. J., & Alpers, G. W. (2014). Emotional pictures and sounds: A review of multimodal interactions of emotion cues in multiple domains. *Frontiers in Psychology*, *5*, 1–13. https://doi.org/10.3389/fpsyg.2014.01351

Guetta, R., & Loui, P. (2017). When music is salty: The crossmodal associations between sound and taste. *PLoS One*, *12*(3), 1–14. https://doi.org/10.1371/journal.pone.0173366

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432–446. https://doi.org/10.1037/xge0000053

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

Kiernan, F., Krause, A. E., & Davidson, J. W. (2022). The impact of biographical information about a composer on emotional responses to their music. *Musicae Scientiae*, *26*(3), 558–584. https://doi.org/10.1177/1029864920988883

Klein, K., Melnyk, V., & Völckner, F. (2021). Effects of background music on evaluations of visual images. *Psychology & Marketing*, *38*(12), 2240–2246. https://doi.org/10.1002/mar.21588

Leder, H., & Pelowski, M. (2021). Empirical aesthetics: Context, extra information, and framing. In M. Nadal & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics* (921–942). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198824350.013.43

Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience Letters*, *455*(2), 129–133. https://doi.org/10.1016/j.neulet.2009.03.044

Margulis, E. H. (2017). An exploratory study of narrative experiences of music. *Music Perception*, *35*(2), 235–248. https://doi.org/10.1525/MP.2017.35.2.235

Margulis, E. H., Wong, P. C. M., Turnbull, C., Kubit, B. M., & McAuley, J. D. (2022). Narratives imagined in response to instrumental music reveal culture-bounded intersubjectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(4), Article e2110406119. https://doi.org/10.1073/pnas.2110406119

Marin, M. M., Gingras, B., & Bhattacharya, J. (2012). Crossmodal transfer of arousal, but not pleasantness, from the musical to the visual domain. *Emotion*, *12*(3), 618–631. https://doi.org/10.1037/a0025020

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Individual differences in music reward experiences. *Music Perception*, *31*(2), 118–138. https://doi.org/10.1525/mp.2013.31.2.118

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: Metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *31*(4), Article A93. https://doi.org/10.1364/josaa.31.000a93

McAuley, J. D., Wong, P. C. M., Mamidipaka, A., Phillips, N., & Margulis, H. (2021). Do you hear what I hear? Perceived narrative constitutes a semantic dimension for music. *Cognition*, *212*, Article 104712. https://doi.org/10.1016/j.cognition.2021.104712

Millis, K. (2001). Making meaning brings pleasure: The influence of titles on aesthetic experiences. *Emotion*, *1*(3), 320–329. https://doi.org/10.1037/1528-3542.1.3.320

Minsky, L., & Fahey, C. (2017). *Audio branding: Using sound to build your brand*. Kogan Page Publishers.

Nunnally, J. C. (1970). *Introduction to psychological measurement*. McGraw-Hill.

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music-color associations are mediated by emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(22), 8836–8841. https://doi.org/10.1073/pnas.1212562110

Pan, F., Zhang, L., Ou, Y., & Zhang, X. (2019). The audio-visual integration effect on music emotion: Behavioral and physiological evidence. *PLoS One*, *14*(5), 1–21. https://doi.org/10.1371/journal.pone.0217040

Parise, C. V., & Spence, C. (2009). “When birds of a feather flock together”: Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One*, *4*(5), Article e5664. https://doi.org/10.1371/journal.pone.0005664

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top–down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews*, *21*, 80–125. https://doi.org/10.1016/j.plrev.2017.02.003

Rosenfeld, N., & Steffens, J. (2019). Effects of audiovisual congruency on perceived emotions in film. *Psychomusicology: Music, Mind, and Brain*, *29*(4), 200–208. https://doi.org/10.1037/pmu0000242

Schlotz, W., Wallot, S., Omigie, D., Masucci, M. D., Hoelzmann, S. C., Vessel, E. A., Wallot, S., & Omigie, D. (2020). The aesthetic responsiveness assessment (AReA): A screening tool to assess individual differences in responsiveness to art in English and German. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(4), 682–696. https://doi.org/10.1037/aca0000348

Schwabe, K., Menzel, C., Mullin, C., Wagemans, J., & Redies, C. (2018). Gist perception of image composition in abstract artworks. *i-Perception*, *9*(3), Article 204166951878079. https://doi.org/10.1177/2041669518780797

Shank, D. B., Stefanik, C., Stuhlsatz, C., Kacirek, K., & Belfi, A. M. (2022). AI Composer bias: Listeners like music less when they think it was composed by an AI. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000447

Turpin, M. H., Walker, A. C., Kara-Yakoubian, M., Gabert, N. N., Fugelsang, J. A., & Stolz, J. A. (2019). Bullshit makes the art grow profounder. *Judgment and Decision Making*, *14*(6), 658–670. https://doi.org/10.2139/ssrn.3410674

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744–756. https://doi.org/10.3758/BF03193776

Wallmark, Z., Nghiem, L., & Marks, L. E. (2021). Does timbre modulate visual perception? Exploring crossmodal interactions. *Music Perception*, *39*(1), 1–20. https://doi.org/10.1525/mp.2021.39.1.1

Whiteford, K. L., Schloss, K. B., Helwig, N. E., & Palmer, S. E. (2018). Color, music, and emotion: Bach to the blues. *i-Perception*, *9*(6), Article 2041669518808535. https://doi.org/10.1177/2041669518808535

Received June 16, 2022
Revision received December 12, 2022
Accepted January 25, 2023 ■